DOCUMENT RESUME

ED 059 633 FL 002 833

AUTHOR Liberman, Alvin M.; And Others
TITLE Language Codes and Memory Codes.
INSTITUTION Haskins Labs., New Haven, Conn.
REPORT NO SR-27-71
PUB DATE 71
NOTE 21p.; In Speech Research, 1 July-30 September 1971, p59-88, Paper presented at meeting on Coding Theory in Learning and Memory, Woods Hole, Massachusetts, August 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Acoustic Phonetics; *Articulation (Speech); Artificial Speech; *Auditory Perception; Distinctive Features; *Grammar; Language; Language Patterns; Language Universals; Linguistic Theory; Memory; Psycholinguistics; *Recall (Psychological); Spectrograms; *Speech; Syllables

ABSTRACT

Paraphrase, as it reflects the processes of remembering rather than those of forgetting, implies that language is best transmitted in one form and stored in another. The dual representation of linguistic information that is implied by paraphrase is important for storing information that has been received and for transmitting information that has been stored. Such duality implies a process of recoding that is somehow constrained by a grammar. Grammar is seen as a set of complex codes that relates transmitted sound and stored meaning. This paper considers the construction characteristics of the speech code which unites the acoustic signal for transmission and the phonetic representation appropriate for storage in short-term memory. The speech code in terms of memory research is then considered with speculation on the construction of the memory code. (Author/VM)

Language Codes and Memory Codes*

Alvin M. Liberman,+ Ignatius G. Mattingly,++ and Michael T. Turvey++
Haskins Laboratories, New Haven

INTRODUCTION: PARAPHRASE, GRAMMATICAL CODES, AND MEMORY

When people recall linguistic information, they commonly produce utterances different in form from those originally presented. Except in special cases where the information does not exceed the immediate memory span, or where rote memory is for some reason required, recall is always a paraphrase.

There are at least two ways in which we can look at paraphrase in memory for linguistic material and linguistic episodes. We can view paraphrase as indicating the considerable degree to which detail is forgotten; at best, what is retained are several choice words with a certain syntactic structure, which, together, serve to guide and constrain subsequent attempts to reconstruct the original form of the information. On this view, rote recall is the ideal, and paraphrase is so much error. Alternatively, we can view the paraphrase not as an index of what has been forgotten but rather as an essential condition or correlate of the processes by which we normally remember. On this view, rote recall is not the ideal, and paraphrase is something other than failure to recall. It is evident that any large amount of linguistic information is not, and cannot be, stored in the form in which it was presented. Indeed, if it were, then we should probably have run out of memory space at a very early age.

We may choose, then, between two views of paraphrase: the first would say that the form of the information undergoes change because of forgetting; the second, that the processes of remembering make such change all but inevitable. In this paper we have adopted the second view, that paraphrase reflects the processes of remembering rather than those of forgetting. Putting this view another way, we should say that the ubiquitous fact of paraphrase implies that language is best transmitted in one form and stored in another.

The dual representation of linguistic information that is implied by paraphrase is important, then, if we are to store information that has been received and to transmit information that has been stored. We take it that such duality implies, in turn, a process of recoding that is somehow

---

*Paper presented at meeting on Coding Theory in Learning and Memory, sponsored by the Committee on Basic Research in Education, Woods Hole, Mass., August 1971.

+Also University of Connecticut, Storrs, and Yale University, New Haven.

++Also University of Connecticut, Storrs.

Aeknowledgments: The authors are indebted for many useful criticisms and suggestions to Franklin S. Cooper of the Haskins Laboratories and Mark Y. Liberman of the United States Army.

constrained by a grammar. Thus, the capacity for paraphrase reflects the fundamental grammatical characteristics of language. We should say, therefore, that efficient memory for linguistic information depends, to a considerable extent, on grammar.

To illustrate this point of view, we might imagine languages that lack a significant number of the grammatical devices that all natural languages have. We should suppose that the possibilities for recoding and paraphrase would, as a consequence, be limited, and that the users of such languages would not remember linguistic information very well. Pidgins appear to be grammatically impoverished and, indeed, to permit little paraphrase, but unfortunately for our purposes, speakers of pidgins also speak some natural language, so they can convert back and forth between the natural language and the pidgin. Sign language of the deaf, on the other hand, might conceivably provide an interesting test. At the present time we know very little about the grammatical characteristics of sign language, but it may prove to have recoding (and hence paraphrase) possibilities that are, by comparison with natural languages, somewhat restricted.[1] If so, one could indeed hope to determine the effects of such restriction on the ability to remember.

In natural languages we cannot explore in that controlled way the causes and consequences of paraphrase, since all such languages must be assumed to be very similar in degree of grammatical complexity. Let us, therefore, learn what we can by looking at the several levels or representations of information that we normally find in language and at the grammatical components that convert between them.

At the one extreme is the acoustic level, where the information is in a form appropriate for transmission. As we shall see, this acoustic representation is not the whole sound as such but rather a pattern of specifiable events, the acoustic cues. By a complexly encoded connection, the acoustic cues reflect the "features" that characterize the articulatory gestures and so the phonetically distinct configurations of the vocal tract. These latter are a full level removed from the sound in the structure of language; when properly combined, they are roughly equivalent to the segments of the phonetic representation.

Only some fifteen or twenty features are needed to describe the phonetics of all human languages (Chomsky and Halle, 1968). Any particular language uses only a dozen or so features from the total ensemble, and at any particular moment in the stream of speech only six or eight features are likely to be significant. The small number of features and the complex relation between sound and feature reflect the properties of the vocal tract and the ear and also, as we will show, the mismatch between these organ systems and the requirements of the phonetic message.

At the other end of the linguistic structure is the semantic representation in which the information is ultimately stored. Because of its relative inaccessibility, we cannot speak with confidence about the shape of the

---

[1]The possibilities for paraphrase in sign language are, in fact, being investigated by Edward Klima and Ursula Bellugi.

information at this level, but we can be sure it is different from the acoustic. We should suppose, as many students do, that the semantic information is also to be described in terms of features. But if the indefinitely many aspects of experience are to be represented, then the available inventory of semantic features must be very large, much larger surely than the dozen or so phonetic features that will be used as the ultimate vehicles. Though particular semantic sets may comprise many features, it is conceivable that the structure of a set might be quite simple. At all events, the characteristics of the semantic representation can be assumed to reflect properties of long-term memory, just as the very different characteristics of the acoustic and phonetic representations reflect the properties of components most directly concerned with transmission.

The gap between the acoustic and semantic levels is bridged by grammar. But the conversion from the one level to the other is not accomplished in a single step, nor is it done in a simple way. Let us illustrate the point with a view of language like the one developed by the generative grammarians (see Chomsky, 1965). On that view there are three levels--deep structure, surface structure, and phonetic representation--in addition to the two--acoustic and semantic--we have already talked about. As in the distinction between acoustic and semantic levels, the information at every level has a different structure. At the level of deep structure, for example, a string such as The man sings. The man married the girl. The girl is pretty. becomes at the surface The man who sings married the pretty girl. The restructuring from one level to the next is governed by the appropriate component of the grammar. Thus, the five levels or streams of information we have identified would be connected by four sets of grammatical rules: from deep structure to the semantic level by the semantic rules; in the other direction, to surface structure, by syntactic rules; then to phonetic representation by phonologic rules; and finally to the acoustic signal by the rules of speech.[2] It should be emphasized that none of these conversions is straightforward or trivial, requiring only the substitution of one segment or representation for another. Nor is it simply a matter of putting segments together to form larger units, as in the organization of words into phrases and sentences or of phonetic segments into syllables and breath groups. Rather, each grammatical conversion is a true restructuring of the information in which the number of segments, and often their order, is changed, sometimes drastically. In the context of the conference for which this paper was prepared, it is appropriate to describe the conversions from one linguistic level to another as recodings and to speak of the grammatical rules which govern them as codes.

Paraphrase of the kind we implied in our opening remarks would presumably occur most freely in the syntactic and semantic codes. But the speech code, at the other end of the linguistic structure, also provides for a kind of paraphrase. At all events it is, as we hope to show, an essential component

---

[2] In generative grammar, as in all others, the conversion between phonetic representation and acoustic signal is not presumed to be grammatical. As we have argued elsewhere, however, and as will to some extent become apparent in this paper, this conversion is a complex recoding, similar in formal characteristics to the recodings of syntax and phonology (Mattingly and Liberman, 1969; Liberman, 1970).

of the process that makes possible the more obvious forms of paraphrase, as well as the efficient memory which they always accompany.

Grammar is, then, a set of complex codes that relates transmitted sound and stored meaning. It also suggests what it is that the recoding processes must somehow accomplish. Looking at these processes from the speaker's viewpoint, we see, for example, that the semantic features must be replaced by phonological features in preparation for transmission. In this conversion an utterance which is, at the semantic level, a single unit comprising many features of meaning becomes, phonologically, a number of units composed of a very few features, the phonologic units and features being in themselves meaningless. Again, the semantic representation of an utterance in coherent discourse will typically contain multiple references to the same topic. This amounts to a kind of redundancy which serves, perhaps, to protect the semantic representation from noise in long-term memory. In the acoustic representation, however, to preserve such repetitions would unduly prolong discourse. To take again the example we used earlier, we do not say The man sings. The man married the girl. The girl is pretty. but rather The man who sings married the pretty girl. The syntactic rules describe the ways in which such redundant references are deleted. At the acoustic and phonetic levels, redundancy of a very different kind may be desirable. Given the long strings of empty elements that exist there, the rules of the phonologic component predict certain lawful phonetic patterns in particular contexts and, by this kind of redundancy, help to keep the phonetic events in their proper order.

But our present knowledge of the grammar does not provide much more than a general framework within which to think about the problem of recoding in memory. It does not, for example, deal directly with the central problem of paraphrase. If a speaker-hearer has gone from sound to meaning by some set of grammatical rules, what is to prevent his going in the opposite direction by the inverse operations, thus producing a rote rendition of the originally presented information? In this connection we should say on behalf of the grammar that it is not an algorithm for automatically recoding in one direction or the other, but rather a description of the relationships that must hold between the semantic representation, at the one end, and the corresponding acoustic representation at the other. To account for paraphrase, we must suppose that the speaker synthesizes the acoustic representation, given the corresponding semantic representation, while the listener must synthesize an approximately equivalent semantic representation, given the corresponding acoustic representation. Because the grammar only constrains these acts of synthesis in very general ways, there is considerable freedom in the actual process of recoding; we assume that such freedom is essential if linguistic information is to be well remembered.

For students of memory, grammatical codes are unsatisfactory in yet another, if closely related, respect: though they may account for an otherwise arbitrary-appearing relation between streams of information at different levels of the linguistic structure, they do not describe the actual processes by which the human being recodes from the one level to the other, nor does the grammarian intend that they should. Indeed, it is an open question whether even the levels that the grammar assumes--for example, deep structure--have counterparts of some kind in the recoding process.

We might do well, then, to concentrate our attention on just one aspect of grammar, the speech code that relates the acoustic and phonetic representations, because we may then avoid some of the difficulties we encounter in the "higher" or "deeper" reaches of the language. The acoustic and phonetic levels have been accessible to psychological (and physiological) experiment, as a result of which we are able to talk about "real" processes and "real" levels, yet the conversion we find there resembles grammatical codes more generally and can be shown, in a functional as well as a formal sense, to be an integral part of language. We will, therefore, examine in some detail the characteristics of the speech code, having in mind that it reflects some of the important characteristics of the broader class of language codes and that it may, therefore, serve well as a basis for comparison with the memory codes we are supposed to be concerned with. It is the more appropriate that we should deal with the speech code because it comprises the conversion from an acoustic signal appropriate for transmission to a phonetic representation appropriate for storage in short-term memory, a process that is itself of some interest to members of this conference.

CHARACTERISTICS OF THE SPEECH CODE

Clarity of the Signal

It is an interesting and important fact about the speech code that the physical signal is a poor one. We can see that this is so by looking at a spectrographic representation of the speech signal like the one in Figure 1. This is a picture of the phrase "to catch pink salmon." As always in a spectrogram, frequency is on the vertical axis, time on the horizontal; relative intensity is represented by the density, or blackness, of the marks. The relatively darker bands are resonances of the vocal tract, the so-called formants. We know that the lowest two or three of these formants contain almost all of the linguistic information; yet, as we can see, the acoustic energy is not narrowly concentrated there but tends rather to be smeared across the spectrum; moreover, there is at least one higher formant at about 3600 cps that never varies and thus carries no linguistic information at all. This is to say that the linguistically important cues constitute a relatively small part of the total physical energy. To appreciate to what extent this is so, we might contrast speech with the printed alphabet, where the important parts of the signal stand out clearly from the background. We might also contrast a spectrogram of the "real" speech of Figure 1 with a "synthetic" spectrogram like the one in Figure 2, which produces intelligible speech though the formants are unnaturally narrow and sharply defined.

In fact, the speech signal is worse than we have so far said or than we can immediately see just by looking at a spectrogram, for, paradoxically, the formants are most indeterminate at precisely those points where the information they carry is most important. It is, we know, the rapid changes in the frequency position of the formants (the formant transitions) that contain the essential cues for most of the consonants. In the case of the stop consonants, these changes occur in 50 msec or less, and they sometimes extend over ranges as great as 600 cps. Such signals scatter energy and are therefore difficult to specify or to track. Moreover, the difficulty is greatest at the point where they begin, though that is the most important part of the transition for the listener who wants to know the phonetic identity of sound.

Spectrogram of "to catch pink salmon," Natural Speech

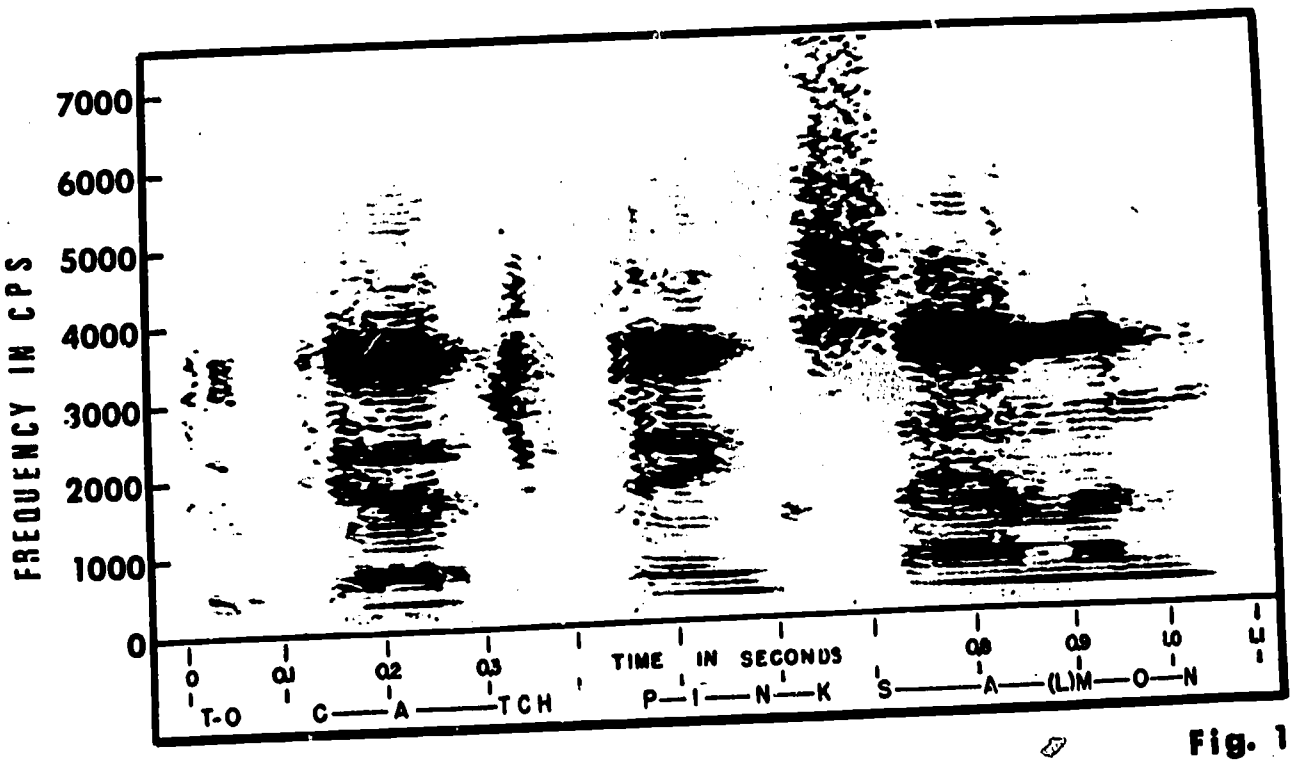

Fig. 1

Schematic Spectrogram for Synthesis of "to catch pink salmon"

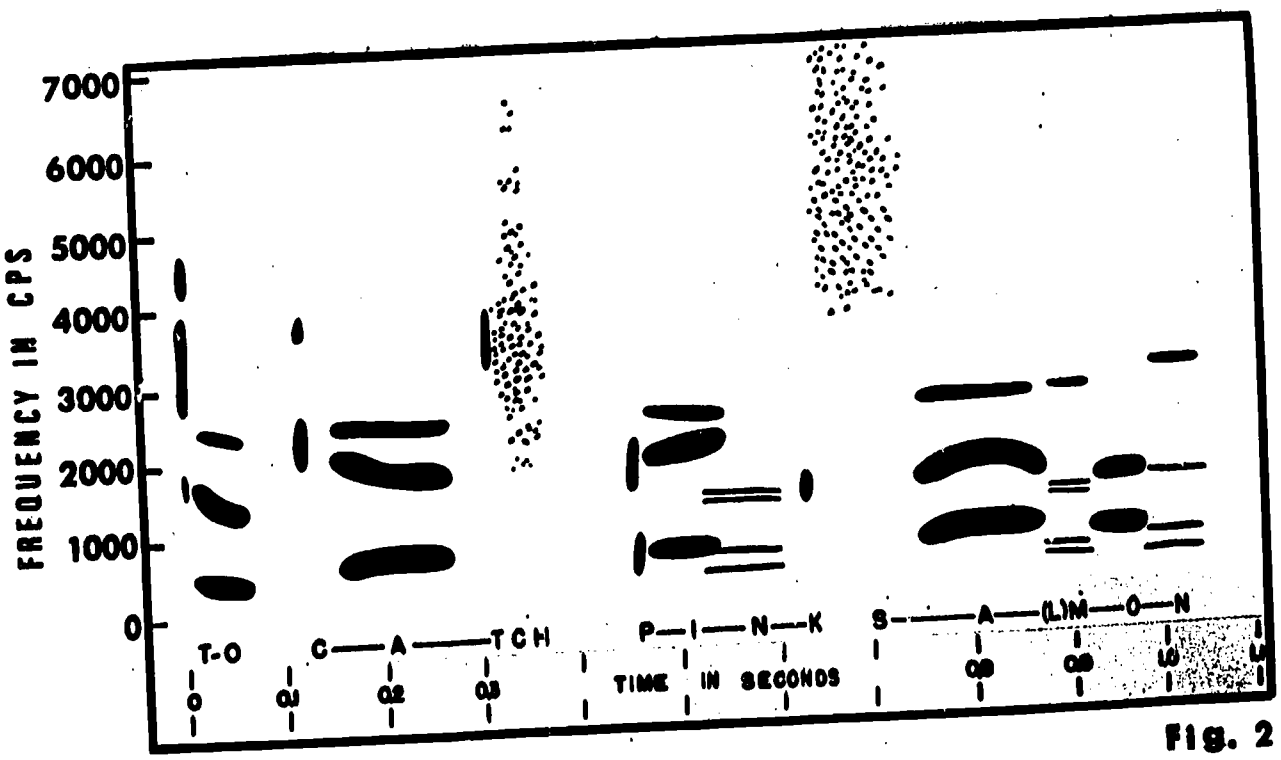

Fig. 2

The physical indeterminacy of the signal is an interesting aspect of the speech code because it implies a need for processors specialized for the purpose of extracting the essential acoustic parameters. The output of these processors might be a cleaned-up description of the signal, not unlike the simplified synthetic spectrogram of Figure 2. But such an output, it is important to understand, would be auditory, not phonetic. The signal would only have been clarified; it would not have been decoded.

Complexity of the Code

Like the other parts of the grammatical code, the conversion from speech sound to phonetic message is complex. Invoking a distinction we have previously found useful in this connection, we should say that the conversion is truly a code and not a cipher (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Studdert-Kennedy, in press). If the sounds of speech were a simple cipher, there would be a unit sound for each phonetic segment. Something approximating such a cipher does indeed exist in one of the written forms of language--viz., alphabets--where each phonological[3] segment is represented by a discrete optical shape. But speech is not an alphabet or cipher in that sense. In the interconversion between acoustic signal and phonetic message the information is radically restructured so that successive segments of the message are carried simultaneously--that is, in parallel--on exactly the same parts of the acoustic signal. As a result, the segmentation of the signal does not correspond to the segmentation of the message; and the part of the acoustic signal that carries information about a particular phonetic segment varies drastically in shape according to context.

In Figure 3 we see schematic spectrograms that produce the syllables [di] and [du] and illustrate several aspects of the speech code. To synthesize the vowels [i] and [u], at least in slow articulation, we need only the steady-state formants--that is, the parts of the pattern to the right of the formant transitions. These acoustic segments correspond in simple fashion to the perceived phonetic segments: they provide sufficient cues for the vowels; they carry information about no other segments; and though the fact is not illustrated here, they are, in slow articulation, the same in all message contexts. For the slowly articulated vowels, then, the relation between sound and message is a simple cipher. The stop consonants, on the other hand, are complexly encoded, even in slow articulation. To see in what sense this is so, we should examine the formant transitions, the rapid changes in formant frequency at the beginning (left) of the pattern. Transitions of the first (lower) formant are cues for manner and voicing; in this case they tell the listener that the consonants are members of the class of voiced stops [bdg]. For our present purposes, the transitions of the second (higher) formant--the parts of the pattern enclosed in the broken circles--are of greater interest. Such transitions are, in general, cues for the perceived "place" distinctions

---

[3]Alphabets commonly make contact with the language at a level somewhat more abstract than the phonetic. Thus, in English the letters often represent what some linguists would call morphophonemes, as for example in the use of "s" for what is phonetically the [s] of cats and the [z] of dogs. In the terminology of generative grammar, the level so represented corresponds roughly to the phonological.

Schematic Spectrogram for the Syllables [di] and [du]

Fig. 3

among the consonants. In the patterns of Figure 3 they tell the listener that the stop is [d] in both cases. Plainly, the transition cues for [d] are very different in the two vowel contexts: the one with [i] is a rising transition relatively high in the spectrum, the one with [u] a falling transition low in the spectrum. It is less obvious, perhaps, but equally true that there is no isolable acoustic segment corresponding to the message segment [d]: at every instant, the second-formant transition carries information about both the consonant and the vowel. This kind of parallel transmission reflects the fact that the consonant is truly encoded into the vowel; this is, we would emphasize, the central characteristic of the speech code.

The next figure (Figure 4) shows more clearly than the last the more complex kind of parallel transmission that frequently occurs in speech. If converted to sound, the schematic spectrogram shown there is sufficient to produce an approximation to the syllable [bæg]. The point of the figure is to show where information about the phonetic segments is to be found in the acoustic signal. Limiting our attention again to the second formant, we see that information about the vowel extends from the beginning of the utterance to the end. This is so because a change in the vowel--from [bæg] to [big], for example--will require a change in the entire formant, not merely somewhere in its middle section. Information about the first consonant, [b], extends through the first two-thirds of the whole temporal extent of the formant. This can be established by showing that a change in the first segment of the message--from [bæg] to [gæg], for example--will require a change in the signal from the beginning of the sound to the point, approximately two-thirds of the way along the formant, that we see marked in the figure. A similar statement and similar test apply also to the last consonant, [g]. In general, every part of the second formant carries information about at least two segments of the message; and there is a part of that formant, in the middle, into which all three message segments have been simultaneously encoded. We see, perhaps more easily than in Figure 1, that the lack of correspondence in segmentation is not trivial. It is not the case that there are simple extensions connecting an otherwise segmented signal, as in the case of cursive writing, or that there are regions of acoustic overlap separating acoustic sections that at some point correspond to the segments of the message. There is no correspondence in segmentation because several segments of the message have been, in a very strict sense, encoded into the same segment of the signal.

Transparency of the Code

We have just seen that not all phonetic segments are necessarily encoded in the speech signal to the same degree. In even the slowest articulations, all of the consonants, except the fricatives,[4] are encoded. But the vowels (and the fricatives) can be, and sometimes are, represented in the acoustic signal quite straightforwardly, one acoustic segment for each phonetic segment. It is as if there were in the speech stream occasionally transparent stretches. We might expect that these stretches, in which the phonetic elements are not restructured in the sound, could be treated as if they were a

---

[4] For a fuller discussion of this point, see Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967.

Schematic Spectrogram Showing Effects of Coarticulation in the Syllable [bæg]

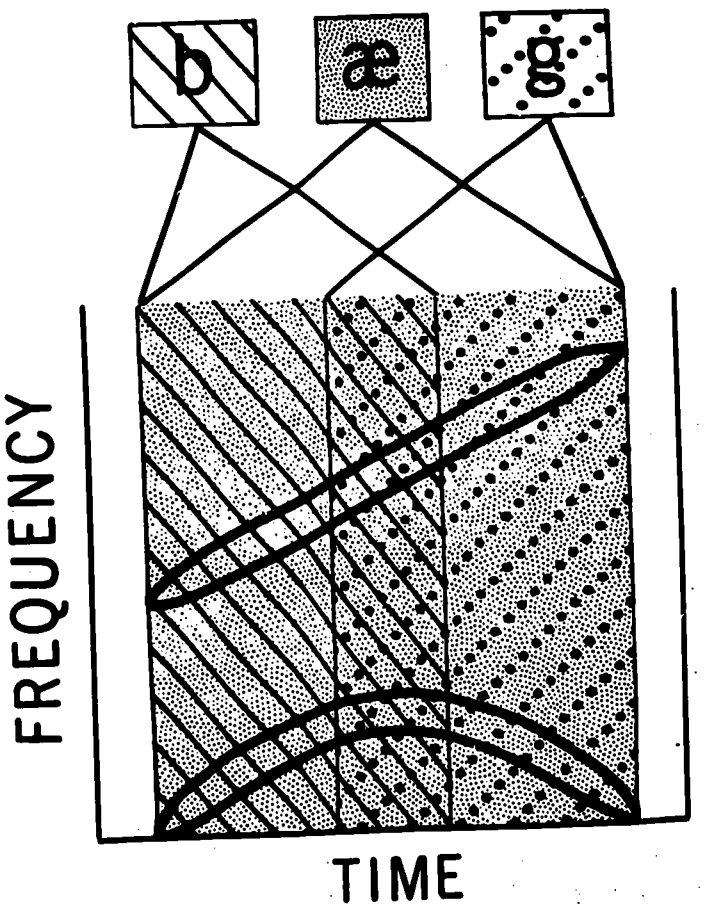

Fig. 4

cipher. There is, thus, a kind of intermittency in the difficulty of decoding the acoustic signal. We may wonder whether that characteristic of the speech code serves a significant purpose--such as providing the decoding machinery with frequent opportunities to get back on the track when and if things go wrong--but it is, in any case, an important characteristic to note, as we will see later in the paper, because of the correspondence between what we might call degree of encoding and evidence for special processing.

Lawfulness of the Code

Given an encoded relation between two streams or levels of information such as we described in the preceding section, we should ask whether the conversion from the one to the other is made lawfully--that is, by the application of rules--or, alternatively, in some purely arbitrary way. To say that the conversion is by rule is to say that it can be rationalized, that there is, in linguistic terms, a grammar. If the connection is arbitrary, then there is, in effect, a code book; to decode a signal, one looks it up in the book.

The speech code is, as we will see, not arbitrary, yet it might appear so to an intelligent but inarticulate cryptanalyst from Mars. Suppose that such a creature, knowing nothing about speech, were given many samples of utterances (in acoustic or visible form), each paired with its decoded or plain-text phonetic equivalents. Let us suppose further, as seems to us quite reasonable, that he would finally conclude that the code could not be rationalized, that it could only be dealt with by reference to a code book. Such a conclusion would, of course, be uninteresting. From the point of view of one who knows that human beings readily decode spoken utterances, the code-book solution would also seem implausible, since the number of entries in the book would have to be so very large. Having in mind the example of [bæg] that we developed earlier, we see that the number of entries would, at the least, be as great as the number of syllables. But, in fact, the number would be very much larger than that, because coding influences sometimes extend across syllable boundaries (Ohman, 1966) and because the acoustic shape of the signal changes drastically with such factors as rate of speaking and phonetic stress (Lindblom, 1963; Lisker and Abramson, 1967).

At all events, our Martian would surely have concluded, to the contrary, that the speech code was lawful if anyone had described for him, even in the most general terms, the processes by which the sounds are produced. Taking the syllable [bæg], which we illustrated earlier, as our example, one might have offered a description about as follows. The phonetic segments of the syllable are taken apart into their constituent features, such as place of production, manner of production, condition of voicing, etc. These features are represented, we must suppose, as neural signals that will become, ultimately, the commands to the muscles of articulation. Before they become the final commands, however, the neural signals are organized so as to produce the greatest possible overlap in activity of the independent muscles to which the separate features are assigned. There may also occur at this stage some reorganization of the commands so as to insure cooperative activity of the several muscle groups, especially when they all act on the same organ, as is the case with the muscle groups that control the gestures of the tongue. But so far the features, or rather their neural equivalents, have only been

organized; they can still be found as largely independent entities, which is to say that they have not yet been thoroughly encoded. In the next stage the neural commands (in the final common paths) cause muscular contraction, but this conversion is, from our standpoint, straightforward and need not detain us. It is in the final conversions, from muscle contraction to vocal-tract shape to sound, that the output is radically restructured and that true encoding occurs. For it is there that the independent but overlapping activity of independent muscle groups becomes merged as they are reflected in the acoustic signal. In the case of [bæg], the movement of the lips that represents a feature of the initial consonant is overlapped with the shaping of the tongue appropriate for the next vowel segment. In the conversion to sound, the number of dimensions is reduced, with the result that the simultaneous activity of lips and tongue affect exactly the same parameter of the acoustic signal, for example, the second formant. We, and our Martian, see then how it is that the consonant and the vowel are encoded.

The foregoing account is intended merely to show that a very crude model can, in general, account for the complexly encoded relation between the speech signal and the phonetic message. That model rationalizes the relation between these two levels of the language, much as the linguists' syntactic model rationalizes the relation between deep and surface structure. For that reason, and because of certain formal similarities we have described elsewhere (Mattingly and Liberman, 1969), we should say of our speech model that it is, like syntax, a grammar. It differs from syntax in that the grammar of speech is a model of a flesh-and-blood process, not, as in the case of syntax, a set of rules with no describable physiological correlates. Because the grammar of speech corresponds to an actual process, we are led to believe that it is important, not just to the scientist who would understand the code but also to the ordinary listener who needs that same kind of understanding, albeit tacitly, if he is to perform appropriately the complex task of perceiving speech. We assume that the listener decodes the speech signal by reference to the grammar, that is, by reference to a general model of the articulatory process. This assumption has been called the motor theory of speech perception.

Efficiency of the Code

The complexity of the speech code is not a fluke of nature that man has somehow got to cope with but is rather an essential condition for the efficiency of speech, both in production and in perception, serving as a necessary link between an acoustic representation appropriate for transmission and a phonetic representation appropriate for storage in short-term memory. Consider production first. As we have already had occasion to say, the constituent features of the phonetic segments are assigned to more or less independent sets of articulators, whose activity is then overlapped to a very great extent. In the most extreme case, all the muscle movements required to communicate the entire syllable would occur simultaneously; in the more usual case, the activity corresponding to the several features is broadly smeared through the syllable. In either case the result is that phonetic segments are realized in articulation at rates higher than the rate at which any single muscle can change its state. The coarticulation that characterizes so much of speech production and causes the complications of the speech code seems well designed to permit relatively slow-moving muscles to transmit phonetic segments at high rates (Cooper, 1966).

The efficiency of the code on the side of perception is equally clear. Consider, first, that the temporal resolving power of the ear must set an upper limit on the rate at which we can perceive successive acoustic events. Beyond that limit the successive sounds merge into a buzz and become unidentifiable. If speech were a cipher on the phonetic message--that is, if each segment of the message were represented by a unit sound--then the limit would be determined directly by the rate at which the phonetic segments were transmitted. But given that the message segments are, in fact, encoded into acoustic segments of roughly syllabic size, the limit is set not by the number of phonetic segments per unit time but by the number of syllables. This represents a considerable gain in the rate at which message segments can be perceived.

The efficient encoding described above results from a kind of parallel transmission in which information about successive segments is transmitted simultaneously on the same part of the signal. We should note that there is another, very different kind of parallel transmission in speech: cues for the features of the same segment are carried simultaneously on <u>different</u> parts of the signal. Recalling the patterns of Figure 4, we note that the cues for place of production are in the second-formant transition, while the first-formant transition carries the cues for manner and voicing. This is an apparently less complicated arrangement than the parallel transmission produced by the encoding of the consonant into the vowel, because it takes advantage of the ear's ability to resolve two very different frequency levels. We should point out, however, that the listener is not at all aware of the two frequency levels, as he is in listening to a chord that is made up of two pitches, but rather hears the stop, with all its features, in a unitary way.

The speech code is apparently designed to increase efficiency in yet another aspect of speech perception: it makes possible a considerable gain in our ability to identify the order in which the message segments occur. Recent research by Warren et al. (1969) has shown that the sequential order of nonspeech signals can be correctly identified only when these segments have durations several times greater than the average that must be assigned to the message segments in speech. If speech were a cipher--that is, if there were an invariant sound for each unit of the message--then it would have to be transmitted at relatively low rates if we were to know that the word "task," for example, was not "taks" or "sakt" or "kats." But in the speech code, the order of the segments is not necessarily signalled, as we might suppose, by the temporal order in which the acoustic cues occur. Recalling what we said earlier about the context-conditioned variation in the cues, we should note now that each acoustic cue is clearly marked by these variations for the position of the signalled segment in the message. In the case of the transition cues for [d] that we described earlier, for example, we should find that in initial and final positions--for example, in [dæg] and [gæd]--the cues were mirror images. In listening to speech we somehow hear through the context-conditioned variation in order to arrive at the canonical form of the segment, in this case [d]. But we might guess that we also use the context-determined shape of the cue to decide where in the sequence the signalled segment occurred. In any case, the order of the segments we hear may be to a large extent inferred--quite exactly synthesized, created, or constructed--from cues in a way that has little or nothing to do with the order of their occurrence in time. Given what appears to be a relatively poor

ability to identify the order of acoustic events from temporal cues, this aspect of the speech code would significantly increase the rate at which we can accurately perceive the message.

The speech code is efficient, too, in that it converts between a high-information-cost acoustic signal appropriate for transmission and a low-information-cost phonetic string appropriate for storage in some short-term memory. Indeed, the difference in information rate between the two levels of the speech code is staggering. To transmit the signal in acoustic form and in high fidelity costs about 70,000 bits per second; for reasonable intelligibility we need about 40,000 bits per second. Assuming a frequency-volley theory of hearing through most of the speech range, we should suppose that a great deal of nervous tissue would have to be devoted to the storage of even relatively short stretches. But recoding into a phonetic representation, we reduce the cost to less than 40 bits per second, thus effecting a saving of about 1,000 times by comparison with the acoustic form and of roughly half that by comparison with what we might assume a reduced auditory (but not phonetic) representation to be. We must emphasize, however, that this large saving is realized only if each phonetic feature is represented by a unitary pattern of nervous activity, one such pattern for each feature, with no additional or extraneous "auditory" information clinging to the edges. As we will see in the next section, the highly encoded aspects of speech do tend to become highly digitized in that sense.

Naturalness of the Code

It is testimony to the naturalness of the speech code that all members of our species acquire it readily and use it with ease. While it is surely true that a child reared in total isolation would not produce phonetically intelligible speech, it is equally true that in normal circumstances he comes to do that without formal tuition. Indeed, given a normal child in a normal environment, it would be difficult to contrive methods that would effectively prevent him from acquiring speech.

It is also relevant that, as we pointed out earlier, there is a universal phonetics. A relatively few phonetic features suffice, given the various combinations into which they are entered, to account for most of the phonetic segments, and in particular those that carry the heaviest information load, in the languages of the world. For example, stops and vowels, the segments with which we have been exclusively concerned in this paper, are universal, as is the co-articulated consonant-vowel syllable that we have used to illustrate the speech code. Such phonetic universals are the more interesting because they often require precise control of articulation; hence they are not to be dismissed with the airy observation that since all men have similar vocal tracts, they can be expected to make similar noises.

Because the speech code is complex but easy, we should suppose that man has access to special devices for encoding and decoding it. There is now a great deal of evidence that such specialized processors do exist in man, apparently by virtue of his membership in the race. As a consequence, speech requires no conscious or special effort; the speech code is well matched to man and is, in precisely that sense, natural.

The existence of special speech processors is strongly suggested by the fact that the encoded sounds of speech are perceived in a special mode. It is obvious--indeed so obvious that everyone takes it for granted--that we do not and cannot hear the encoded parts of the speech signal in auditory terms. The first segment of the syllables [ba], [da], [ga] have no identifiable auditory characteristics; they are unique linguistic events. It is as if they were the abstract output of a device specialized to extract them, and only them, from the acoustic signal. This abstract nonauditory perception is characteristic of encoded speech, not of a class of acoustic events such as the second-formant transitions that are sufficient to distinguish [ba], [da], [ga], for when these transition cues are extracted from synthetic speech patterns and presented alone, they sound just like the "chirps" or glissandi that auditory psychophysics would lead us to expect. Nor is this abstract perception characteristic of the relatively unencoded parts of the speech signal: the steady-state noises of the fricatives, [s] and [ʃ], for example, can be heard as noises; moreover, one can easily judge that the noise of [s] is higher in pitch than the noise of [ʃ].

A corollary characteristic of this kind of abstract perception, measured quite carefully by a variety of techniques, is one that has been called "categorical perception" (see Studdert-Kennedy, Liberman, Harris, and Cooper, 1970, for a review; Haggard, 1970, 1971b; Pisoni, 1971; Vinegrad, 1970). In listening to the encoded segments of speech we tend to hear them only as categories, not as a perceived continuum that can be more or less arbitrarily divided into regions. This occurs even when, with synthetic speech, we produce stimuli that lie at intermediate points along the acoustic continuum that contains the relevant cues. In its extreme form, which is rather closely approximated in the case of the stops, categorical perception creates a situation, very different from the usual psychophysical case, in which the listener can discriminate stimuli as different no better than he can identify them absolutely.

That the categorical perception of the stops is not simply a characteristic of the way we process a certain class of acoustic stimuli--in this case the rapid frequency modulation that constitutes the (second-formant transition) acoustic cue--has been shown in a recent study (Mattingly, Liberman, Syrdal, and Halwes, 1971). It was found there that, when listened to in isolation, the second-formant transitions--the chirps we referred to earlier--are not perceived categorically.

Nor can it be said that categorical perception is simply a consequence of our tendency to attach phonetic labels to the elements of speech and then to forget what the elements sounded like. If that were the case, we should expect to find categorical perception of the unencoded steady-state vowels, but in fact, we do not--certainly not to the same extent (Fry, Abramson, Eimas, and Liberman, 1962; Eimas, 1963; Stevens, Liberman, Ohman, and Studdert-Kennedy, 1969; Pisoni, 1971; Fujisaki and Kawashima, 1969). Moreover, categorical perception of the encoded segments has recently been found to be reflected within 100 msec in cortical evoked potentials (Dorman, 1971).

In the case of the encoded stops, then, it appears that the listener has no auditory image of the signal available to him, but only the output of a specialized processor that has stripped the signal of all normal sensory

information and represented each phonetic segment (or feature) categorically by a unitary neural event. Such unitary neural representations would presumably be easy to store and also to combine, permute, and otherwise shuffle around in the further processing that converts between sound and meaning.

But perception of vowels is, as we noted, not so nearly categorical. The listener discriminates many more stimuli than he can absolutely identify, just as he does with nonspeech; accordingly, we should suppose that, as with nonspeech, he hears the signal in auditory terms. Such an auditory image would be important in the perception of the pitch and duration cues that figure in the prosodic aspects of speech; moreover, it would be essential that the auditory image be held for some seconds, since the listener must often wait to the end of a phrase or sentence in order to know what linguistic value to assign to the particular pitch and duration cues he heard earlier.

Finally, we should note about categorical perception that, according to a recent study (Eimas, Siqueland, Jusczyk, and Vigorito, 1971), it is present in infants at the age of four weeks. These infants discriminated synthetic [ba] and [pa]; moreover, and more significantly, they discriminated better, other things being equal, between pairs of stimuli which straddled the adult phonetic boundary than between pairs which lay entirely within the phonetic category. In other words, the infants perceived the voicing feature categorically. From this we should conclude that the voicing feature is real, not only physiologically but in a very natural sense.

Other, perhaps more direct, evidence for the existence of specialized speech processors comes from a number of recent experiments that overload perceptual mechanisms by putting competing signals simultaneously into the two ears (Broadbent and Gregory, 1964; Bryden, 1963; Kimura, 1961, 1964, 1967; Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). The general finding with speech signals, including nonsense syllables that differ, say, only in the initial consonant, is that stimuli presented to the right ear are better heard than those presented to the left; with complex nonspeech sounds the opposite result--a left-ear advantage--is found. Since there is reason to believe, especially in the case of competing and dichotically presented stimuli, that the contralateral cerebral representation is the stronger, these results have been taken to mean that speech, including its purely phonetic aspects, needs to be processed in the left hemisphere, nonspeech in the right. The fact that phonetic perception goes on in a particular part of the brain is surely consistent with the view that it is carried out by a special processor.

The case for a special processor to decode speech is considerably strengthened by the finding that the right-ear advantage depends on the encodedness of the signal. For example, stop consonants typically show a larger and more consistent right-ear advantage than unencoded vowels (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). Other recent studies have confirmed that finding and have explored even more analytically the conditions of the right-ear (left-hemisphere) advantage for speech (Darwin, 1969, 1971; Haggard, 1971a; Haggard, Ambler, and Callow, 1969; Haggard and Parkinson, 1971; Kirstein and Shankweiler, 1969; Spellacy and Blumstein, 1970). The results, which are too numerous and complicated to present here even in summary form, tend to support the conclusion that processing is forced into

the left hemisphere (for most subjects) when phonetic decoding, as contrasted with phonetic deciphering or with processing of nonspeech, must be carried out.

Having referred in the discussion of categorical perception to the evidence that the phonetic segments (or, rather, their features) may be assumed to be represented by unitary neural events, we should here point to an incidental result of the dichotic experiments that is very relevant to that assumption. In three experiments (Halwes, 1969; Studdert-Kennedy and Shankweiler, 1970; Yoder, pers. comm.) it has been found that listeners tend significantly often to extract one feature (e.g., place of production) from the input to one ear and another feature (e.g., voicing) from the other and combine them to hear a segment that was not presented to either ear. Thus, given [ba] to the left ear, say, and [ka] to the right, listeners will, when they err, far more often report [pa] (place feature from the left ear, voicing from the right) or [ga] (place feature from the right ear, voicing from the left) than [da] or [ta]. We take this as conclusive evidence that the features are singular and unitary in the sense that they are independent of the context in which they occur and also that, far from being abstract inventions of the linguist, they have, in fact, a hard reality in physiological and psychological processes.

The technique of overloading the perceptual machinery by dichotic presentation has led to the discovery of yet another effect which seems, so far, to testify to the existence of a special speech processor (Studdert-Kennedy, Shankweiler, and Schulman, 1970). The finding, a kind of backward masking that has been called the "lag" effect, is that when syllables contrasting in the initial stop consonant are presented dichotically and offset in time, the second (or lagging) syllable is more accurately perceived. When such syllables are presented monotically, the first (or leading) stimulus has the advantage. In the dichotic case, the effect is surely central; in the monotic case there is presumably a large peripheral component. At all events, it is now known that, as in the case of the right-ear advantage, the lag effect is greater for the encoded stops than for the unencoded vowels (Kirstein, 1971; Porter, Shankweiler, and Liberman, 1969); it has also been found that highly encoded stops show a more consistent effect than the relatively less encoded liquids and semi-vowels (Porter, 1971). Also relevant is the finding that synthetic stops that differ only in the second-formant transitions show a lag effect but that the second-formant transitions alone (that is, the chirps) do not (Porter, 1971). Such results support the conclusion that this effect, too, may be specific to the special processing of speech.[5]

In sum, there is now a great deal of evidence to support the assertion that man has ready access to physiological devices that are specialized for the purpose of decoding the speech signal and recovering the phonetic message. Those devices make it possible for the human being to deal with the speech code easily and without conscious awareness of the process or its complexity. The code is thus a natural one.

---

[5]One experimental result appears so far not to fit with that conclusion: syllables that differed in a linguistically irrelevant pitch contour nevertheless gave a lag effect (Darwin, in press).

Resistance to Distortion

Everyone who has ever worked with speech knows that the signal holds up well against various kinds of distortion. In the case of sentences, a great deal of this resistance depends on syntactic and semantic constraints, which are, of course, irrelevant to our concern here. But in the perception of nonsense syllables, too, the message often survives attempts to perturb it. This is due largely to the presence in the signal of several kinds of redundancy. One arises from the phonotactic rules of the language: not all sequences of speech sounds are allowable. That constraint is presumably owing, though only in part, to limitations having to do with the possibilities of co-articulation. In any case, it introduces redundancy and may serve as an error-correcting device. The other kind of redundancy arises from the fact that most phonetic distinctions are cued by more than one acoustic difference. Perception of place of production of the stop consonants, for example, is normally determined by transitions of the second formant, by transitions of the third formant, and by the frequency position of a burst of noise. Each of these cues is more or less sufficient, and they are highly independent of each other. If one is wiped out, the others remain.

There is one other way in which speech resists distortion that may be the most interesting of all because it implies for speech a special biological status. We refer here to the fact that speech remains intelligible even when it is removed about as completely as it can be from its normal, naturalistic context. In the synthetic patterns so much used by us and others, we can, and often do, play fast and loose with the nature of the vocal-tract excitation and with such normally fixed characteristics of the formants as their number, bandwidth, and relative intensity. Such departures from the norm, resulting in the most extreme cases in highly schematic representations, remain intelligible. These patterns are more than mere cartoons, since certain specific cues must be retained. As Mattingly (in this Status Report) has pointed out, speech might be said in this respect to be like the sign stimuli that the ethologist talks about. Quite crude and unnatural models such as Tinbergen's (1951) dummy sticklebacks, elicit responses provided only that the model preserves the significant characters of the original display. As Manning (1969:39) says,"sign stimuli will usually be involved where it is important never to miss making a response to the stimulus." More generally, sign stimuli are often found when the correct transmission of information is crucial for the survival of the individual or the species. Speech may have been used in this way by early man.

How to Tell Speech from Nonspeech

For anyone who uses the speech code, and especially for the very young child who is in the process of acquiring it, it is necessary to distinguish the sounds of speech from other acoustic stimuli. How does he do this? The easy, and probably wrong, answer is that he listens for certain acoustic stigmata that mark the speech signal. One thinks, for example, of the nature of the vocal-tract excitation or of certain general characteristics of the formants. If the listener could identify speech on the basis of such relatively fixed markers, he would presumably decide at a low level of the perceptual system whether a particular signal was speech or not and, on the basis of that decision, send it to the appropriate processors. But we saw in the

preceding section that speech remains speech even when the signal is reduced to an extremely schematic form. We suspect, therefore, that the distinction between speech and nonspeech is not made at some early stage on the basis of general acoustic characteristics.

More compelling support for that suspicion is to be found in a recent experiment by T. Rand (pers. comm.) To one ear he presented all of the first formant, including the transitions, together with the steady-state parts of the second and third formants; when presented alone, these patterns sound vaguely like [da]. To the other ear, with proper time relationships carefully preserved, were presented the 50-msec second-formant and third-formant transitions; alone, these sound like the chirps we have referred to before. But when these patterns were presented together--that is, dichotically--listeners clearly heard [ba], [da] or [ga] (depending on the nature of the second-formant and third-formant transitions) in one ear and, simultaneously, nonspeech chirps in the other. Thus, it appears that the same acoustic events--the second-formant or third-formant transitions--can be processed simultaneously as speech and nonspeech. We should suppose, then, that the incoming signal goes indiscriminately to speech and nonspeech processors. If the speech processors succeed in extracting phonetic features, then the signal is speech; if they fail, then the signal is processed only as nonspeech. We wonder if this is a characteristic of all so-called sign stimuli.

Security of the Code

The speech code is available to all members of the human race, but probably to no other species. There is now evidence that animals other than man, including even his nearest primate relatives, do not produce phonetic strings and their encoded acoustic correlates (Lieberman, 1968, 1971; Lieberman, Klatt, and Wilson, 1969; Lieberman, Crelin, and Klatt, in press). This is due, at least in part, to gross differences in vocal-tract anatomy between man and all other animals. (It is clear that speech in man is not simply an overlaid function, carried out by peripheral structures that evolved in connection with other more fundamental biological processes; rather, some important characteristics of the human vocal tract must be supposed to have developed in evolution specifically in connection with speech.) Presumably, animals other than man lack also the mechanisms of neurological control necessary for the organization and coordination of the gestures of speech, but hard evidence for this is lacking. Unfortunately, we know nothing at all about how animals other than man perceive speech. Presumably, they lack the special processor necessary to decode the speech signal. If so, we must suppose that their perception of speech would be different from ours. They should not hear categorically, for instance, and they should not hear the [di]-[du] patterns of Figure 3 as two-segment syllables which have the first segment in common. Thus, we should suppose that animals other than man can neither produce nor correctly perceive the speech code. If all our enemies were animals other than man, cryptanalysts would have nothing to do--or else they might have the excessively difficult task of breaking an animal code for which man has no natural key.

Subcodes

Our discussion so far has, perhaps, left the impression that there is only one speech code. In one sense this is true, for it appears that there

is a universal ensemble of phonetic features defined by the communicative possibilities of the vocal tract and the neural speech processor. But the subset of phonetic features which are actually used varies from language to language. Each language thus has its own phonetic "subcode." A given phonetic feature, however, will be articulated and perceived in the same way in every language in which it is used. Thus, we should be very surprised, for instance, to find a language in which the perception of place for stops was not categorical. If, as Eimas's results lead us to suppose, a child is born with an intuitive knowledge of the universal phonetics, part of his task in learning his native language is to identify the features of its phonetic subcode and to forget the others. These unused features cannot be entirely lost, however, since people do learn how to speak and understand more than one language. But there is some evidence that bilinguals listening to their second language do not necessarily use the same speech cues as native speakers of the language do (Haggard, 1971b).

Secondary Codes

A speaker-hearer can become aware of certain aspects of the linguistic process, in particular its phonological and phonetic processes. The awareness can then be exploited to develop "secondary codes," which may be thought of as additional pseudolinguistic rules added to those of the language. A simple example is a children's "secret language," such as Pig Latin, in which a rule for metathesis and insertion applies to each word. We should suppose that to speak or understand Pig Latin fluently would require not only the unconscious knowledge of the linguistic structure of English that all native speakers have, but also a conscious awareness of a particular aspect of this structure--the phonological segmentation--and a considerable amount of practice. There is evidence, indeed, that speakers of English who lack a conscious awareness of phonological segmentation do not master Pig Latin, despite the triviality of its rules (Savin, in press). The pseudolinguistic character of Pig Latin explains why even a speaker of English who does not know Pig Latin would not mistake it for a natural foreign language, and why one continues to feel a sense of artificiality in speaking it long after he has mastered the trick.

Systems of versification are more important kinds of secondary codes. For a literate society the function of verse is primarily esthetic, but for preliterate societies, verse is a means of transmitting verbal information of cultural importance with a minimum of paraphrase. The rules of verse are, in effect, an addition to the phonology which requires that recalled material not only should preserve the semantic values of the original, but should also conform to a specific, rule-determined phonetic pattern. Thus in Latin epic poetry, a line of verse is divided into six feet, each of which must have one of several patterns of long and short syllables. The requirement to conform to this pattern excludes almost all possible renditions other than the correct one and makes memorization easier and recall more accurate. Since versification rules are in general more elaborate than those of Pig Latin, a greater degree of linguistic awareness is necessary to compose verse. This more complex skill has thus traditionally been the specialized occupation of a few members of a society, though a passive form of the skill, permitting the listener to distinguish"correct" from "incorrect" lines without scanning them syllable by syllable, has been possible for a much larger number of people.

Writing, like versification, is also a secondary code for transmitting verbal information accurately, and the two activities have more in common than might at first appear. The reader is given a visually coded representation of the message, and this representation, whether ideographic, syllabic, or alphabetic, provides very incomplete information about the linguistic structure and semantic content of the message. The skilled reader, however, does not need complete information and ordinarily does not even need all of the partial information given by the graphic patterns but rather just enough to exclude most of the other messages which might fit the context. Being competent in his language, knowing the rules of the writing system, and having some degree of linguistic awareness, he can reproduce the writer's message in reasonably faithful fashion. (Since the specific awareness required is awareness of phonological segmentation, it is not surprising that Savin's group of English speakers who cannot learn Pig Latin also have great difficulty in learning to read.)

The reader's reproduction is not, as a rule, verbatim; he makes small deviations which are acceptable paraphrases of the original and overlooks or, better, unconsciously corrects misprints. This suggests that reading is an active process of construction constrained by the partial information on the printed page, just as remembering verse is an active process of construction, constrained, though much less narrowly, by the rules of versification. As Bartlett (1932) noted for the more general case, the processes of perception and recall of verbal material are not essentially different.

For our purposes, the significant fact about pseudolinguistic secondary codes is that, while being less natural than the grammatical codes of language, they are nevertheless far from being wholly unnatural. They are more or less artificial systems based on those aspects of natural linguistic activities which can most readily be brought to consciousness: the levels of phonology and phonetics. All children do not acquire secondary codes maturationally, but every society contains some individuals who, if given the opportunity, can develop sufficient linguistic awareness to learn them, just as every society has its potential dancers, musicians, and mathematicians.

LANGUAGE, SPEECH, AND RESEARCH ON MEMORY

What we have said about the speech code may be relevant to research on memory in two ways: most directly, because work on memory for linguistic information, to which we shall presently turn, naturally includes the speech code as one stage of processing; and, rather indirectly, because the characteristics of the speech code provide an interesting basis for comparison with the kinds of code that students of memory, including the members of this conference, talk about. In this section of the paper we will develop that relevance, summarizing where necessary the appropriate parts of the earlier discussion.

The Speech Code in Memory Research

Acoustic, auditory, and phonetic representations. When a psychologist deals with memory for language, especially when the information is presented as speech sounds, he would do well to distinguish the several different forms that the information can take, even while it remains in the domain of speech. There is, first, the acoustic form in which the signal is transmitted. This

is characterized by a poor signal-to-noise ratio and a very high bit rate. The second form, found at an early stage of processing in the nervous system, is auditory. This neural representation of the information maps in a relatively straightforward way onto the acoustic signal. Of course, the acoustic and auditory forms are not identical. In addition to the fact that one is mechanical and the other neural, it is surely true that some information has been lost in the conversion. Moreover, as we pointed out earlier in the paper, it is likely that the signal has been sharpened and clarified in certain ways. If so, we should assume that the task was carried out by devices not unlike the feature detectors the neurophysiologist and psychologist now investigate and that apparently operate in visual perception, as they do in hearing, to increase contrast and extract certain components of the pattern. But we should emphasize that the conversion from acoustic to auditory form, even when done by the kind of device we just assumed, does not decode the signal, however much it may improve it. The relation of the auditory to the acoustic form remains simple, and the bit rate, though conceivably a good deal lower at this neural stage than in the sound itself, is still very high. To arrive at the phonetic representation, the third form that the information takes, requires the specialized decoding processes we talked about earlier in the paper. The result of that decoding is a small number of unitary neural patterns, corresponding to phonetic features, that combine to make the somewhat greater number of patterns that constitute the phonetic segments; arranged in their proper order, these segments become the message conveyed by the speech code. The phonetic representations are, of course, far more economical in terms of bits than the auditory ones. They also appear to have special standing as unitary physiological and biological realities. In general, then, they are well suited for storage in some kind of short-term memory until enough have accumulated to be recoded once more, with what we must suppose is a further gain in economy.

Even when language is presented orthographically to the subjects' eyes, the information seems to be recoded into phonetic form. One of the most recent and also most interesting treatments of this matter is to be found in a paper by Conrad (in press). He concludes, on the basis of considerable evidence, that while it is possible to hold the alphabetic shapes as visual information in short-term memory--deaf-mute children seem to do just that--the information can be stored (and dealt with) more efficiently in phonetic form. We suppose that this is so because the representations of the phonetic segments are quite naturally available in the nervous system in a way, and in a form, that representations of the various alphabetic shapes are not. Given the complexities of the conversion from acoustic or auditory form to phonetic, and the advantages for storage of the phonetic segments, we should insist that this is an important distinction.

Storage and transmission in man and machine. We have emphasized that in spoken language the information must be in one form (acoustic) for transmission and in a very different form (phonetic or semantic) for storage, and that the conversion from the one to the other is a complex recoding. But there is no logical requirement that this be so. If all the components of the language system had been designed from scratch and with the same end in view, the complex speech code might have been unnecessary. Suppose the designer had decided to make do with a smaller number of empty segments, like the phones we have

been talking about, that have to be transmitted in rapid succession. The engineer might then have built articulators able to produce such sequences simply--alphabetically or by a cipher--and ears that could perceive them. Or if he had, for some reason, started with sluggish articulators and an ear that could not resolve rapid-fire sequences of discrete acoustic signals, he might have used a larger inventory of segments transmitted at a lower rate. In either case the information would not have had to be restructured in order to make it differentially suitable for transmission and storage; there might have been, at most, a trivial conversion by means of a simple cipher. Indeed, that is very much the situation when computers "talk" to each other. The fact that the human being cannot behave so simply, but must rather use a complex code to convert between transmitted sound and stored message, reflects the conflicting design features of components that presumably developed separately and in connection with different biological functions. As we noted in an earlier part of the paper, certain structures, such as the vocal tract, that evolved originally in connection with nonlinguistic functions have undergone important modifications that are clearly related to speech. But these adaptations apparently go only so far as to make possible the further matching of components brought about by devices such as those that underlie the speech code.

It is obvious enough that the ear involved long before speech made its appearance, so we are not surprised, when we approach the problem from that point of view, to discover that not all of its characteristics are ideally suited to the perception of speech. But when we consider speech production and find that certain design features do not mesh with the characteristics of the ear, we are led to wonder if there are not aspects of the process--in particular, those closer to the semantic and cognitive levels--that had independently reached a high state of evolutionary development before the appearance of language as such and had then to be imposed on the best available components to make a smoothly functioning system. Indeed, Mattingly (this Status Report) has explicitly proposed that language has two sources, an intellect capable of semantic representation and a system of "social releasers" consisting of articulated sounds, and that grammar evolved as an interface between these two very different mechanisms.

In the alphabet, man has invented a transmission vehicle for language far simpler than speech--a secondary code, in the sense discussed earlier. It is a straightforward cipher on the phonological structure, one optical shape for each phonological segment, and has a superb signal-to-noice ratio. We should suppose that it is precisely the kind of transmission vehicle that an engineer might have devised. That alphabetic representations are, indeed, good engineering solutions is shown by the relative ease with which engineers have been able to build the so-called optical character readers. However, the simple arrangements that are so easy for machines can be hard for human beings. Reading comes late in the child's development; it must be taught; and many fail to learn. Speech, on the other hand, bears a complex relation to language as we have seen and has so far defeated the best efforts of engineers to build a device that will perceive it. Yet this complex code is mastered by children at an early age, some significant proficiency being present at four weeks; it requires no tuition; and everyone who can hear manages to perceive speech quite well.

The relevance of all this to the psychology of memory is an obvious and generally observed caution: namely, that we be careful about explaining human beings in terms of processes and concepts that work well in intelligent and remembering machines. We nevertheless make the point because we have in speech a telling object lesson. The speech code is an extremely complex contrivance, apparently designed to make the best of a bad fit between the requirement that phonetic segments be transmitted at a rapid rate and the inability of the mouth and the ear to meet that requirement in any simple way. Yet the physiological devices that correct this mismatch are so much a part of our being that speech works more easily and naturally for human beings than any other arrangement, including those that are clearly simpler.

More and less encoded elements of speech. In describing the characteristics of the speech code we several times pointed to differences between stop consonants and vowels. The basic difference has to do with the relation between signal and message: stop consonants are always highly encoded in production, so their perception requires a decoding process; vowels can be, and sometimes are, represented by encipherment, as it were alphabetically, in the speech signal, so they might be perceived in a different and simpler way. We are not surprised, then, that stops and vowels differ in their tendencies toward categorical perception as they do also in the magnitude of the right-ear advantage and the lag effect (see above).

An implication of this characteristic of the speech code for research in immediate memory has appeared in a study by Crowder (in press) which suggests that vowels produce a "recency" effect, but stops do not. Crowder and Morton (1969) had found that, if a list of spoken words is presented to a subject, there is an improvement in recall for the last few items on the list, but no such recency effect is found if the list is presented visually. To explain this model difference, Crowder and Morton suggested that the spoken items are held for several seconds in an "echoic" register in "precategorical" or raw sensory form. At the time of recall these items are still available to the subject in all their original sensory richness and are therefore easily remembered. When presented visually, the items are held in an "iconic" store for only a fraction of a second. In his more recent experiment Crowder has found that for lists of stop-vowel syllables, the auditory recency effect appears if the syllables on the list contrast only in their vowels but is absent if they contrast only in their stops. If Crowder and Morton's interpretation of their 1969 result is correct, at least in general terms, then the difference in recency effect between stops and vowels is exactly what we should expect. As we have seen in this paper, the special process that decodes the stops strips away all auditory information and presents to immediate perception a categorical linguistic event the listener can be aware of only as [b,d,g,p,t, or k]. Thus, there is for these segments no auditory, precategorical form that is available to consciousness for a time long enough to produce a recency effect. The relatively unencoded vowels, on the other hand, are capable of being perceived in a different way. Perception is more nearly continuous than categorical: the listener can make relatively fine discriminations within phonetic classes because the auditory characteristics of the signal can be preserved for a while. (For a relevant model and supporting data see Fujisaki and Kawashima, 1969.) In the experiment by Crowder, we may suppose that these same auditory characteristics of the vowel, held

for several seconds in an echoic sensory register, provide the subject with the rich, precategorical information that enables him to recall the most recently presented items with relative ease.

It is characteristic of the speech code, and indeed of language in general, that not all elements are psychologically and physiologically equivalent. Some (e.g., the stops) are more deeply linguistic than others (e.g., the vowels); they require special processing and can be expected to behave in different ways when memory codes are used.

Speech as a special process. Much of what we said about the speech code was to show that it is complex in a special way and that it is normally processed by a correspondingly special device. When we examine the formal aspects of this code, we see resemblances of various kinds to the other grammatical codes of phonology and syntax--which is to say that speech is an integral part of a larger system called language--but we do not readily find parallels in other kinds of perception. We know very little about how the speech processor works, so we cannot compare it very directly with other kinds of processors that the human being presumably uses. But knowing that the task it must do appears to be different in important ways from the tasks that confront other processors, and knowing, too, that the speech processor is in one part of the brain while nonspeech processors are in another, we should assume that speech processing may be different from other kinds. We might suppose, therefore, that the mechanisms underlying memory for linguistic information may be different from those used in other kinds of memory such as, for example, visual or spatial.

Speech appears to be specialized, not only by comparison with other perceptual or cognitive systems of the human being, but also by comparison with any of the systems so far found in other animals. While there may be some question about just how many of the so-called higher cognitive and linguistic processes monkeys are capable of, it seems beyond dispute that the speech code is unique to man. To the extent, then, that this code is used in memory processes--for example, in short-term memory--we must be careful about generalizing results across species.

Speech and Memory Codes Compared

It will be recalled that we began by adopting the view that paraphrase has more to do with the processes by which we remember than with those by which we forget. In this vein we proposed that when people are presented with long stretches of sensible language, they normally use the devices of grammar to recode the information from the form in which it was transmitted into a form suitable for storage. On the occasion of recall they code it back into another transmittable form that may resemble the input only in meaning. Thus, grammar becomes an essential part of normal memory processes and of the memory codes that this conference is about. We therefore directed our attention to grammatical codes, taking these to be the rules by which conversions are carried out from one linguistic level to another. To spell out the essential features of such codes, we chose to deal in detail with just one, the speech code. It can be argued, persuasively we think, that the speech code is similiar to other grammatical codes, so its characteristics can be used, within reasonable limits, to represent those of grammar generally. But

speech has the advantage in this connection that it has been more accessible to psychological investigation than the other grammatical codes. As a result, there are experimental data that permit us to characterize speech in ways that provide a useful basis for comparison with the codes that have come from the more conventional research on verbal memory. In this final section we turn our attention briefly to those more conventional memory codes and to a comparison between them and the speech code.

We will apply the same convention to this discussion of conventional memory codes that we applied to our discussion of grammatical codes. That is, the term "code" is reserved for the rules which convert from one representation of the information to another. In our analysis of the speech code we took the acoustic and phonetic levels as our two representations and inferred the properties of the speech code from the relation between the two.

In the most familiar type of experiment the materials the subject is required to remember are not the longer segments of language, such as sentences or discourses, but rather lists of words or nonsense syllables. Typically in such an experiment, the subject is required to reproduce the information exactly as it was presented to him, and his response is counted as an error if he does not. Under those circumstances it is difficult, if not impossible, for the subject to employ his linguistic coding devices to their fullest extent, or in their most normal way. However, it is quite evident that the subject in this situation nevertheless uses codes; moreover, he uses them for the same general purpose to which, we have argued, language is so often put, which is to enable him to store the information in a form different from that in which it was presented. Given the task of remembering unfamiliar sequences such as consonant trigraphs, the subject may employ, sometimes to the experimenter's chagrin, some form of linguistic mediation (Montague, Adams, and Kiess, 1966). That is, he converts the consonant sequence into a sentence or proposition, which he then stores along with a rule for future recovery of the consonant string. In a recent examination of how people remember nonsense syllables, Prytulak (1971) concluded that such mediation is the rule rather than the exception. Reviewing the literature on memory for verbal materials, Tulving and Madigan (1970) describe two kinds of conversions: one is the substitution of an alternative symbol for the input stimulus together with a conversion rule; the other is the storage of ancillary information along with the to-be-remembered item. Most generally, it appears that when a subject is required to remember exactly lists of unrelated words, paired-associates, or digit strings, he tries to impart pattern to the material, to restructure it in terms of familiar relationships. Or he resorts, at least in some situations, to the kind of "chunking" that Miller (1956) first described and that has become a staple of memory theory (Mandler, 1967). Or he converts the verbal items into visual images (Paivio, 1969; Bower, 1970). At all events, we find that, as Bower (1970) has pointed out, bare-bones rote memorization is tried only as a last resort, if at all.

The subject converts to-be-remembered material which is unrelated and relatively meaningless into an interconnected, meaningful sequence of verbal items or images for storage. What can be said about the rules relating the two levels? In particular, how do the conversions between the two levels compare with those that occur in the speech code, and thus, indirectly, in

language in general? The differences would appear to be greater than the similarities. Many of these conversions that we have cited are more properly described as simple ciphers than as codes, in the sense that we have used these terms earlier, since there is in these cases no restructuring of the information but only a rather straightforward substitution of one representation for another. Moreover, memory codes of this type are arbitrary and idiosyncratic, the connection between the two forms of the information having arisen often out of the accidents of the subject's life history; such rules as there may be (for example, to convert each letter of the consonant trigraph to a word beginning with that letter) do not truly rationalize the code but rather fall back, in the end, on a key that is, in effect, a code book. As often as not, the memory codes are also relatively unnatural: they require conscious effort and, on occasion, are felt by the subject to be difficult and demanding. In regard to efficiency, it is hard to make a comparison; relatively arbitrary and unnatural codes can nevertheless be highly efficient given enough practice and the right combination of skills in the user.

In memory experiments which permit the kind of remembering characterized by paraphrase, we would expect to find that memory codes would be much like language codes, and we should expect them to have characteristics similar to those of the code we know as speech. The conversions would be complex recodings, not simple substitutions; they would be capable of being rationalized; and they would, of course, be highly efficient for the uses to which they were being put. But we would probably find their most obvious characteristic to be that of naturalness. People do not ordinarily contrive mnemonic aids by which to remember the gist of conversations or of books, nor do they necessarily devise elaborate schemes for recalling stories and the like, yet they are reasonably adept at such things. They remember without making an effort to commit a message to memory; more important, they do not have to be taught how to do this sort of remembering.

It is, of course, exceedingly difficult to do scientific work in situations that permit the free use of these very natural language codes. Proper controls and measures are hard to arrange. Worse yet, the kinds of paraphrase that inevitably occur in long discourses will span many sentences and imply recoding processes so complex that we hardly know now how to talk about them. Yet, if the arbitrary, idiosyncratic ciphers which we have described are simply devices to mold to-be-remembered, unrelated materials into a form amenable to the natural codes, then it must be argued that our understanding of such ciphers will advance more surely with knowledge of the natural bases from which they derive and to which they must, presumably, be anchored.

REFERENCES

Bartlett, F.C. (1932) Remembering. (Cambridge, England: Cambridge University Press).

Bower, G.H. (1970) Organizational factors in memory. Cog. Psychol. 1, 18-46.

Broadbent, D.E. and Gregory, M. (1964) Accuracy of recognition for speech presented to the right and left ears. Quart. J. exp. Psychol. 16, 359-360.

Bryden, M.P. (1963) Ear preference in auditory perception. J. exp. Psychol. 65, 103-105.

Chomsky, N. (1965) Aspects of the Theory of Syntax. (Cambridge, Mass.: M.I.T. Press).

Chomsky, N. and Halle, M. (1968) The Sound Pattern of English. (New York: Harper and Row).

Conrad, R. (in press) Speech and reading. In Language by Ear and by Eye: The Relationships between Speech and Reading, J.F. Kavanagh and I.G. Mattingly, eds. (Cambridge, Mass.: M.I.T. Press).

Cooper, F.S. (1966) Describing the speech process in motor command terms. J. acoust. Soc. Amer. 39, 1221A. (Text in Haskins Laboratories Status Report on Speech Research SR-5/6, 1966.)

Crowder, R. (in press) The sound of vowels and consonants in immediate memory. J. verb. Learn. verb Behav., 10.

Crowder, R.B. and Morton, J. (1969) Precategorical and acoustic storage (PAS). Perception and Psychophysics 5, 365-373.

Darwin, C.J. (1969) Auditory Perception and Cerebral Dominance. Unpublished doctoral dissertation, University of Cambridge.

Darwin, C.J. (1971) Ear differences in the recall of fricatives and vowels. Quart. J. exp. Psychol. 23, 46-62.

Darwin, C.J. (in press) Dichotic backward masking of complex sounds. Quart. J. exp. Psychol.

Dorman, M. (1971) Auditory Evoked Potential Correlates of Speech Perception. Unpublished doctoral dissertation, University of Connecticut.

Eimas, P.D. (1963) The relation between identification and discrimination along speech and nonspeech continua. Language and Speech 3, 206-217.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., and Vigorito, J. (1971) Speech perception in infants. Science 171, 303-306.

Fry, D.B., Abramson, A.S., Eimas, P.D. and Liberman, A.M. (1962) The identification and discrimination of synthetic vowels. Language and Speech 5, 171-189.

Fujisaki, H. and Kawashima, T. (1969) On the modes and mechanisms of speech perception. In Annual Report No. 1. (Tokyo: University of Tokyo, Division of Electrical Engineering, Engineering Research Institute).

Haggard, M.P. (1970) Theoretical issues in speech perception. In Speech Synthesis and Perception 4. (Cambridge, England: Psychological Laboratory).

Haggard, M.P. (1971a) Encoding and the REA for speech signals. Quart. J. exp. Psychol. 23, 34-45.

Haggard, M.P. (1971b) New demonstrations of categorical perception. In Speech Synthesis and Perception 5. (Cambridge, England: Psychological Laboratory).

Haggard, M.P., Ambler, S. and Callow, M. (1969) Pitch as a voicing cue. J. acoust. Soc. Amer. 47, 613-617.

Haggard, M.P. and Parkinson, A.M. (1971) Stimulus and task factors as determinants of ear advantages. Quart. J. exp. Psychol. 23, 168-177.

Halwes, T. (1969) Effects of Dichotic Fusion on the Perception of Speech. Unpublished doctoral dissertation, University of Minnesota. (Reproduced as Supplement to Haskins Laboratories Status Report on Speech Research 1969.)

Kimura, D. (1961) Cerebral dominance and perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura D. (1964) Left-right differences in the perception of melodies. Quart. J. exp. Psychol. 16, 355-358.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Kirstein, E. (1971) Temporal Factors in the Perception of Dichotically Presented Stop Consonants and Vowels. Unpublished doctoral dissertation, University of Connecticut. (Reproduced in Haskins Laboratories Status Report on Speech Research SR-24.)

Kirstein, E. and Shankweiler, D.P. (1969) Selective listening for dichotically presented consonants and vowels. Paper read before 40th Annual Meeting of Eastern Psychological Association, Philadelphia, 1969. (Text in Haskins Laboratories Status Report on Speech Research SR-17/18, 133-141.)

Liberman, A.M. (1970) The grammars of speech and language. Cog. Psychol. 1, 301-323.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Lieberman, P. (1968) Primate vocalizations and human linguistic ability. J. acoust. Soc. Amer. 44, 1574-1584.

Lieberman, P. (1971) On the speech of Neanderthal man. Linguistic Inquiry 2, 203-222.

Lieberman, P., Klatt, D., and Wilson, W.A. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates. Science 164, 1185-1187.

Lieberman, P., Crelin, E.S., and Klatt, D.H. (in press) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. American Anthropologist. (Also in Haskins Laboratories Status Report on Speech Research SR-24, 51-90.)

Lindblom, B. (1963) Spectrographic study of vowel reduction. J. acoust. Soc. Amer. 35, 1773-1781.

Lisker, L. and Abramson, A.S. (1967) Some effects of context on voice onset time in English stops. Language and Speech 10, 1-28.

Mandler, G. (1967) Organization and memory. In The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 1, K.W. Spence and J.T. Spence, eds. (New York: Academic Press).

Manning, A. (1969) An Introduction to Animal Behavior. (Reading, Mass.: Addison-Wesley).

Mattingly, I.G. (This Status Report) Speech cues and sign stimuli.

Mattingly, I.G. and Liberman, A.M. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, K.N. Leibovic, ed. (New York: Springer Verlag).

Mattingly, I.G., Liberman, A.M., Syrdal, A.K., and Halwes, T. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

Miller, G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev. 63, 81-97.

Montague, W.E., Adams, J.A., and Kiess, H.O. (1966) Forgetting and natural language mediation. J. exp. Psychol. 72, 829-833.

Ohman, S.E.G. (1966) Coarticulation in VCV utterances: Spectrographic measurements. J. acoust. Soc. Amer. 39, 151-168.

Paivio, A. (1969) Mental imagery in associative learning and memory. Psychol. Rev. 76, 241-263.

Pisoni, D. (1971) On the Nature of Categorical Perception of Speech Sounds. Unpublished doctoral dissertation, University of Michigan. (Reproduced as Supplement to Haskins Laboratories Status Report on Speech Research, 1971.)

Porter, R.J. (1971) Effects of a Delayed Channel on the Perception of Dichotically Presented Speech and Nonspeech Sounds. Unpublished doctoral dissertation, University of Connecticut.

Porter, R., Shankweiler, D.P., and Liberman, A.M. (1969) Differential effects of binaural time differences in perception of stop consonants and vowels. Paper presented at annual meeting of the American Psychological Association, Washington, D.C., 2 September.

Prytulak, L.S. (1971) Natural language mediation. Cog. Psychol. 2, 1-56.
Savin, H. (in press) What the child knows about speech when he starts learning to read. In Language by Ear and by Eye: The Relationships between Speech and Reading, J.F. Kavanagh and I.G. Mattingly, eds. (Cambridge, Mass.: M.I.T. Press).
Shankweiler, D. and Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. exp. Psychol. 19, 59-63.
Spellacy, F. and Blumstein, S. (1970) The influence of language set on ear preference in phoneme recognition. Cortex 6, 430-439.
Stevens, K.N., Liberman, A.M., Ohman, S.E.G., and Studdert-Kennedy, M. (1969) Cross-language study of vowel perception. Language and Speech 12, 1-23.
Studdert-Kennedy, M. (in press) The perception of speech. In Current Trends in Linguistics, T.A. Sebeok, ed. (The Hague: Mouton). (Also in Haskins Laboratories Status Report on Speech Research SR-23, 15-48.)
Studdert-Kennedy, M., Liberman, A.M., Harris, K.S., and Cooper, F.S. (1970) Motor theory of speech perception: A reply to Lane's critical review. Psychol. Rev. 77, 234-249.
Studdert-Kennedy, M. and Shankweiler, D. (1970) Hemispheric specialization for speech perception. J. acoust. Soc. Amer. 48, 579-594.
Studdert-Kennedy, M., Shankweiler, D., and Schulman, S. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. acoust. Soc. Amer. 48, 599-602.
Tinbergen, N. (1951) The Study of Instinct. (Oxford: Clarendon Press).
Tulving, E. and Madigan, S.A. (1970) Memory and verbal learning. Annual Rev. Psychol. 21, 437-484.
Vinegrad, M. (1970) A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. Haskins Laboratories Status Report on Speech Research SR-21/22, 147-156.
Warren, R.M., Obusek, C.J., Farmer, R.M., and Warren, R.T. (1969) Auditory sequence: Confusions of patterns other than speech or music. Science 164, 586-587.